LAMP-TR-040 April 2000
UMIACS-TR-2000-17
CS-TR-4120

Large-Scale Construction of a Chinese-English Semantic Hierarchy

Bonnie J. Dorr, Gina-Anne Levow, Dekang Lin

Language and Media Processing Labratory
Instititue for Advanced Computer Studies
College Park, MD 20742

Abstract

This paper addresses the problem of building conceptual resources for multilingual applications. We describe new techniques for large-scale construction of a semantic hierarchy for Chinese verbs, using thematic-role information lto create links between Chinese concepts and English classes. We then present an approach to compensating for gaps in the existing resources. The resulting hierarchy is used for a multilingual lexicon for Chinese-English machine translation and cross-language information retrieval applications.

***The support of the LAMP Technical Report Series and the partial support of this research by the National Science Foundation under grant EIA0130422 and the Department of Defense under contract MDA9049-C6-1250 is gratefully acknowledged.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
APR 2000		**00-04-2000 to 00-04-2000**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Large-Scale Construction of a Chinese-English Semantic Hierarchy	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Language and Media Processing Laboratory,Institute for Advanced Computer Studies,University of Maryland,College Park,MD,20742-3275**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**13**	

Large-Scale Construction of Chinese-English Semantic Hierarchy

Bonnie Dorr, Gina Levow, and Dekang Lin

University of Maryland and University of Manitoba

{bonnie,gina,lindek}@umiacs.umd.edu

Abstract: This paper describes an approach to large-scale construction of a semantic hierarchy for Chinese verbs. Leveraging off of an existing Chinese conceptual database called HowNet and a Levin-based English verb classification, we use thematic-role information to create links between Chinese concepts and English classes. The resulting hierarchy is used for multilingual lexicons in machine translation and cross-language information retrieval applications.

1 Introduction

The growing quantity of online multilingual information has created an urgent need for rapid construction of lexical resources. Automatic and semi-automatic techniques for lexical acquisition are more critical now than ever before as it becomes infeasible to produce adequate semantic representations on a large scale by human labor alone.

We describe an approach to large-scale construction of a semantic hierarchy for Chinese verbs. Leveraging off of an existing classification of English verbs called EVCA (English Verbs Classes and Alternations) (Levin, 1993) and a Chinese conceptual database called HowNet (Zhendong, 1988c; Zhendong, 1988b; Zhendong, 1988a) (http://www.how-net.com), we use thematic-role information (e.g., a mapping between the HowNet "Patient" and the EVCA-based "Th(eme)") to create links between Chinese concepts and English classes. Each Chinese-English link is additionally associated with a sense from WordNet (Miller and Fellbaum, 1991), thus producing a new Asian companion to the current (Euro)WordNet initiative.

We focus on the assignment of word senses to semantically classified verbs. The resulting lexicons are used

for determining appropriate word senses in applications such as machine translation and cross-language information retrieval. The importance of word-sense disambiguation to either of these two applications is clear when one considers the degree of inaccuracy that might result from using a weak alternative, such access to a bilingual word list.

For example, the Chinese verb 拉 (la) corresponds to a wide range of English glosses—even if we examine only the verb translations—in the Optilex Chinese-English dictionary: *slash*, *cut*, *chat*, *pull*, *drag*, *transport*, *move*, *raise*, *help*, *implicate*, *involve*, *defecate*, *pressgang*.[1] Our work provides a framework for disambiguating such cases in a given context by associating certain of these senses (e.g., *transport*, *move*) with one HowNet concept (e.g., |Transport|) while associating other senses (e.g., *help*) to another HowNet concept (e.g., |help|).

Several researchers have investigated the problem of assigning class-based senses to verbs (Dorr, 1997), (Palmer and Rosenzweig, 1996), (Palmer and Wu, 1995) using a variety of online resources including Longman's Dictionary of Contemporary English (LDOCE) (Procter, 1978), EVCA (Levin, 1993), and WordNet (Miller and Fellbaum, 1991). The work of (Nomura et al., 1994), (Saint-Dizier, 1996), (Jones et al., 1994) indicates that the translation of English classes into other languages is not straightforward, but later work has shown that regularities between different language classifications is evident in online resources (Dang et al., 1998), (Dorr and Jones, 1999), (Olsen et al., 1998).

This work extends the techniques described by (Palmer and Wu, 1995), which used a concept space to produce a hierarchical organization of Chinese verbs. The extensions include: (1) The use of the entire EVCA database rather than a small set of verbs (the *break* class); (2) The provision of a thematic-role based filter for a more refined version of verb-class assignments. Later work by (Dang et al., 1998) uses an intersective-class technique that partitions English verbs into refined classes using WordNet as a conceptual basis. We adopt a technique that is similar in flavor to this approach, with the following extensions: (1) Concept alignment across two different language hierarchies (Chinese and English) rather than one; (2)

[1] Optilex is a large (600k entries) machine readable Chinese-English dictionary; although this dictionary is in some ways exhaustive, there is no encoding of part-of-speech information, but see (Olsen et al., 1998) for a description of a procedure that extracts verbs automatically from Optilex.

Hooks into WordNet senses for both languages; (3) Mappings between Chinese and English thematic roles.

The EVCA classes used in this work relies on extensions by (Dorr, 1997) and (Dorr and Jones, 1996) to a finer-grained set of semantic classes, including 26 new classes. There are 485 total classes in the extended set, each hand-tagged with WordNet senses and thematic-role specifications. Mapping English roles to their Chinese counterparts is the primary aid in associating WordNet senses with Chinese verbs; the thematic-role mappings are used as a guideline for selecting the appropriate entry in EVCA, which in turn is associated with a WordNet sense.

We will demonstrate that it is possible to produce a lexicon by associating 478 Chinese HowNet concepts with 485 EVCA classes, with a clear concept-to-class correspondence in a large majority of the cases. We will describe how this correspondence is extracted and we will show how this process has provided a framework for compensating for gaps in our online resources. The lexicon resulting from this approach is large-scale, containing 17284 Chinese-English conceptual links.

2 Mapping Between Chinese HowNet and English EVCA

The mapping between Chinese HowNet and English EVCA involves three steps:

(1) Produce all possible English Optilex *glosses* (translations) for all 12342 Chinese verbs in HowNet and associate each Chinese verb with one or more of the 478 HowNet concepts—forming 48,884 verb-to-concept candidates.

Example: The multiply ambiguous Chinese verb 拉 (la) has several different Optilex glosses (*slash*, *cut*, *chat*, *pull*, *drag*, *transport*, *move*, *raise*, *help*, *implicate*, *involve*, *defecate*, *pressgang*) and is associated with multiple HowNet concepts: |Transport|, |Attract|, |Excrete|, |Force|, |Help|, |Include|, |Pull|, |Recreation|, and |Talk|.

(2) Associate each verb-to-concept candidate with one or more of the 485 EVCA classes—forming an average of 2 thousand verb-to-class entries per HowNet concept (on the order of 1 million verb-to-class

candidates, total).

Example: The Chinese verb 拉 (la) is associated with 22 EVCA classes: Admire (31.2.b, *implicate*, *involve*); Amuse (31.1.b, *transport*, *move*, *cut*; Braid (41.2.2, *cut*); Breathe (40.1.2, *defecate*); Build (26.1.a, *cut*); Carry (11.4.i, *carry*, *pull*, *drag*); Chitchat (37.6.a, *chat*); Crane (40.3.2, *raise*); Cut (21.1.a, *slash*, *cut*); Cut (21.1.d, *cut*); Equip (13.4.2, *help*); Force (12.a.ii, *pull*); Get (13.5.1.a, *pull*); Grow (26.2.a.ii, *raise*); Hurt (40.8.3, *pull*, *cut*); Meander (47.7.a, *cut*); Play (009, *pawn*); Put (9.4.a, *raise*); Search (35.2.a, *drag*); Send (11.1, *smuggle*, *transport*, *ship*, *convey*); Send Slide (11.2.b, *move*); Split (23.2.b, *cut*, *pull*).

(3) For each HowNet concept, partition the associated Chinese-English pairs into groups whose English glosses correspond EVCA classes. This requires three steps:

   a. Order the candidate EVCA classes so that the highest-ranking classes are those that contain the highest number of English verbs matching the Optilex glosses.

   b. In cases where a tie-breaker is needed, reorder the candidate EVCA classes according to the degree to which the thematic-role specification in HowNet concept matches that of EVCA class. The matching procedure relies on the correlations shown in Table 1 which were derived from approximately 200 seed mappings.[2]

   c. For each Chinese-English entry associated with the HowNet concept, assign the highest ranking candidate EVCA class.

Example: Two of the HowNet concepts associated with the multiply ambiguous Chinese verb 拉 (la) are |Help| and |Transport|. The thematic-role specification associated with |Help| is `(agent,patient,scope)` (as in *John helped him with his work*). This specification most closely matches that of Equip EVCA Class (where 拉 (la) is translated as *help*) which has the specification `_ag_th,mod-poss(with)`; thus, the |Help| HowNet concept is associated with the Equip EVCA Class, and the mapping between the

[2]The seed mappings were done by hand at a rate of approximately 50 mappings per hour; these were verified and by a native Chinese speaker in a half day.

two is `(agent->ag),(patient->th),(scope->mod-poss)`.[3]

On the other hand, the |Transport| HowNet concept is associated with the thematic-role specification `(agent,patient,LocationIni,LocationFin,direction)` (as in *John transported the goods from Boston to New York (westward)*). This specification most closely matches that of the Send EVCA Class (where 拉 (la) is translated as *transport*); thus, the |Transport| Hownet concept is associated with the Send EVCA class, and the mapping between the two is `(agent->ag),(patient->th),(LocationIni->src),(LocationFin->goal)`.

The end result is that the English glosses associated with 拉 (la) are filtered down to *help* in the Equip semantic class and *transport* in the Send semantic class; the corresponding WordNet senses are assigned (for free) from the hand-tagged EVCA database. These are Senses 1–3 in the case of *transport* (i.e., *move/carry/displace*) and Sense 1 in the case of *help* (i.e., *aid/assist*).

The process of associating EVCA classes with Chinese verbs relies on a massive filtering of spurious class assignments. For example, the |Establish| HowNet concept is ultimately associated with only two EVCA classes, 29.2.c and 26.4.a (Characterize and Create), but it initially had 29 potential EVCA class assignments. One example of an EVCA class that was ruled out is the Change of State class, 45.4.a, associated with the Optilex translation *colonize* for the Chinese verb 殖民. (zhimin) Although this is a perfectly valid EVCA class assignment for the HowNet concept |Colonize|, it is not appropriate for the |Establish| HowNet concept. Because this class is ranked 8th for |Establish|—as opposed to 1st and 2nd place ranking for 29.2.c and 26.4.a, respectively—this assignment is ruled out by our algorithm.

3 Results

The histogram in Table 2 characterizes the number of EVCA classes required for coverage of 478 HowNet concepts. We consider the approach to be a success for several reasons: (1) Association of a unique EVCA

[3]Thematic-role specifications and their use in generation of natural-language translations are described in (Dorr et al., 1998).

Hownet	EVCA Semantic Roles														
Semantic Roles	**ag**	**th**	**exp**	**goal**	**src**	**perc**	**loc**	**info**	**pred**	**prop**	**Instr**	**Poss**	**Pred**	**Purp**	**Loc**
agent	278	77	32	1	2	3	0	0	0	0	4	7	0	11	0
beneficiary	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cause	0	0	0	1	0	4	0	0	1	6	4	7	1	11	0
content	0	31	1	2	2	14	0	20	3	6	3	0	1	3	0
contrast	0	2	0	1	0	1	0	0	0	0	0	1	0	0	0
cost	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
degree	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
direction	0		0	3	0	0	5	0	0	0	0	0	0	0	2
duration	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
experiencer	13	32	33	0	0	0	0	0	0	0	0	0	0	0	0
instrument	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
isa	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0
location	0	1	0	1	0	0	6	0	0	1	2	0	0	0	2
manner	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
material	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0
partner	0	2	0	0	3	3	0	0	0	0	0	11	0	0	0
partof	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
patient	0	122	7	7	0	8	0	0	0	0	0	0	0	0	0
possession	0	28	0	0	1	2	0	0	0	0	0	3	0	0	0
purpose	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
range	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
relevant	15	4	4	0	0	1	1	0	0	0	0	0	0	0	0
result	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
scope	0	1	0	0	0	2	1	0	0	0	1	2	3	0	0
source	0	4	0	0	16	0	0	0	0	0	0	0	0	0	3
target	0	7	12	27	1	17	0	0	0	3	0	2	0	0	1
time	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
whole	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
ContentProduct	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0
DurationAfterEvent	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
LocationFin	0	0	0	31	0	0	8	0	1	0	0	2	2	0	8
LocationIni	0	0	0	0	24	0	2	0	0	0	0	0	0	0	0
PartOfTouch	0	0	0	0	1	0	1	0	0	1	1	1	0	0	2
PatientProduct	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0
ResultEvent	0	0	0	5	0	0	0	3	0	7	1	1	2	0	0
ResultIsa	0	0	0	0	1	0	0	0	2	1	0	0	1	0	0
ResultWhole	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1
SourceWhole	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
StateFin	0	0	0	5	0	1	0	0	0	0	0	0	0	0	0
StateIni	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0
TimeFin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TimeIni	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 1: Seed Table for mapping HowNet Roles into EVCA Roles

Number of EVCA Classes per Concept:	0	1	2	3	4	5
Number of HowNet Concepts:	2	371	71	20	10	4

Table 2: Histogram of HowNet Concept Partitions into EVCA Classes

HowNet Concept	**EVCA Class(es)**
|Transport|	11.1 Send
|Help|	13.4.2 Equip
|Apologize|	32.2.a Long
|Naming|	29.3 Dub
|Judge|	29.4 Declare
|Moisten|	45.4.a Change of State
|Excrete|	40.1.2 Breathe
|TakeVehicle|	51.4.2.a.ii Motion by Vehicle
|PlayDown|	33.b Judgment (75%), 31.2.a Admire (25%)
|Establish|	29.2.c Characterize (90%), 26.4.a Create (19%)
|Decorate|	9.8.b Fill (50%), 26.1.b Build (43%), 9.9.ii Butter (25%)
|Buy|	10.5 Steal (08%), 13.5.1.a Get (30%), 13.5.1.b.ii Get (54%), 13.5.2.d Get (46%)
|Teach|	29.2.c Characterize (24%), 33.b Judgment (71%), 37.9.a Advise (29%), 37.1.a Transfer Message (45%), 31.1.a Amuse (19%)

Table 3: Examples of HowNet Partitionings with Respect to EVCA

class to a HowNet concept was achieved in 371 cases—77% of the HowNet classes; (2) Most of the other cases partitioned the HowNet entries into 2 EVCA classes; (3) Only 2 cases did not correspond to any EVCA class (i.e., every word associated with the concept belonged to a different EVCA class); (4) There were no partitionings exceeding 5 EVCA classes.

Examples of the HowNet partitionings into EVCA classes are given in Table 3, with a focus on the cases where 1 partition was found. In cases where there is more than 1 partition, percentages are given with respect to the number of Chinese verbs in each HowNet class.[4]

4 Compensating for Resource Deficiencies

As part of our effort to produce a complete alignment between HowNet and EVCA, we built an EVCA-based canonical entry for each of the 478 HowNet concepts so that we could compensate for certain types of

[4]The astute reader will notice the percentages don't always total 100%. This is because certain of the Chinese verbs are assigned to two different "partitionings." The resulting groups are, thus, not *true* partitions in the mathematical sense since they are not necessarily mutually exclusive. In the cases where the percentages total 100%, the resulting groups are mutually exclusive.

HowNet Concept	Canonical Entry
|Transport|	11.1 Send, *transport*
|Help|	13.4.2 Equip, *help*
|Apologize|	32.2.a Long, *apologize*
|Naming|	29.3 Dub, *name*
|Judge|	29.4 Declare, *judge*
|Moisten|	45.4.a Change of State, *moisten*
|Excrete|	40.1.2 Breathe, *excrete*
|TakeVehicle|	51.4.2.a.ii Motion by Vehicle, *ride*
|PlayDown|	33.b Judgment, *belittle*
|Establish|	29.2.c Characterize, *establish*
|Decorate|	9.8.b Fill, *decorate*
|Buy|	13.5.1.b.ii, *buy*
|Teach|	37.1.a Transfer Message, *teach*

Table 4: Sample of Canonical Entries for Filling Resource Gaps

resource deficiencies. The canonical entry is specified as an EVCA class coupled with its associated prototype verb. This entry was automatically generated according to the highest ranking EVCA class using steps 3.a and 3.b in Section 2. Each canonical entry was hand-verified (at a rate of 80 per hour for 478 classes). In most cases, prototype word names the HowNet concept, e.g., *transport* for the |Transport| HowNet concept. In other cases—where the HowNet concept is not an English word—the prototype word is a realization of that concept, e.g., *belittle* for the |PlayDown| HowNet concept. A sample of the canonical entries is given in Table 4.

We use these canonical entries to compensate for any gaps that arise in our three online resources: (1) EVCA, (2) Optilex, and (3) HowNet. We will describe each of these, in turn.

4.1 EVCA Gaps

An EVCA gap is detected when an Optilex verb does not occur in EVCA. When this occurs, the canonical entry is automatically used as the appropriate EVCA classification for the verb. For example, one Optilex gloss associated with HowNet concept |Establish| (for the verb 重建 (chongjian)) is *reconstruct*, which does not occur in EVCA. This is a case where the canonical entry (29.2.c Characterize, *establish*) is associated with the verb.

An interesting byproduct of the handling of EVCA gaps is that it allows us to enhance our EVCA resource. For example the verb *reconstruct* can now be added to Class 29.2.c and the WordNet sense associated with the verb *establish* can then be linked to this Chinese verb.

4.2 Optilex Gaps

An Optilex gap occurs when a particular translation for a Chinese verb is missing. For example, in Optilex 摆布 has only one Optilex gloss: *manipulate*. However, the word 摆布 (baibu) is associated with two HowNet concepts, |Decorate| and |Control|. This gloss is only appropriate for the |Control| concept. The *decorate* meaning of 摆布 (baibu) is omitted in Optilex.

Such gaps are detected by means of two types of information: (1) HowNet and EVCA semantic-role specifications; (2) correlations between the gloss under question and *other* HowNet concepts. In this particular example, the semantic-role specification for *manipulate* in EVCA is `(ag,exp,instr)`, which is ranked low (11th out of 28) with respect to the HowNet specification `(agent,patient)` in the |Decorate| class. By contrast, this same EVCA class has a high ranking (2nd out of 22) in the |Control| concept due to a close match between `(ag,exp,instr)` and the HowNet semantic-role specification `(agent,patient,ResultEvent)`. In addition, the correlation of the gloss *manipulate* is much higher for the |Control| concept than it is for the |Decorate| concept (4 occurrences compared to 0). From these two types of information, we can conclude that the *decorate* sense of 摆布 (baibu) is missing from Optilex. As in the case with EVCA gaps, the canonical entry (9.8.b Fill, *decorate*) is associated with the Chinese verb to compensate for this Optilex gap.

In addition to their usefulness in handling of gaps in our lexical resources, the canonical entries proved useful for assigning EVCA classes to Chinese verbs whose Optilex gloss was not "parsable" by our gloss extraction procedure. For example, the Chinese verb 挨打 (aida) has only a single Optilex translation: *take a beating*. This verb is associated with the HowNet concept |Suffer|, which has as its canonical entry (31.3.d Marvel, *suffer*). Thus, the canonical entry was assigned to this verb.

A similar approach is used for unknown or misspelled words. For example, the translation of 输送 (shusong)

as in Optilex is misspelled as *tranport*. Because this verb occurs in the |Transport| class, the canonical entry (11.1 Send, *transport*) was assigned to this verb.

4.3 HowNet Gaps

In some cases, the HowNet classification incorrectly associates a Chinese word with a particular concept. For example, HowNet incorrectly associates the two Chinese verbs 扎花 (zhahua) and 绣花 (xiuhua) with |Decorate|. These two verbs are translated as *embroider* in EVCA class 26.1.b (Build), but their meaning is closer to *sew flowers*. That is, the patient is incorporated into the verb, which means the semantic-role specification `_ag_th_goal(into),ben(for)` does not match that of the HowNet concept `(agent,possession,source)`.

Discrepancies in HowNet are detected by means of frequency within the class. Out of the 17 entries associated with the |Decorate| concept, only two of them (the two misclassed Chinese verbs) are associated with an EVCA class that is not 9.9 or 9.8. As in the gap-recovery described approaches above, the misclassed verbs are associated with the canonical entry (9.8.b Fill, *decorate*).[5]

5 Summary and Future Work

We have presented an approach to aligning two large-scale online resources, HowNet and EVCA. The lexicon resulting from this approach is large-scale, containing 17284 Chinese-English conceptual links. The technique for producing these links involves matching semantic-role specifications in HowNet with those in EVCA. Our results indicate that the correspondence is very high between the 478 Chinese HowNet concepts and the 485 EVCA classes. Because each Chinese-English link is additionally associated with a WordNet sense, we see this resource as the first step toward producing a new Asian language companion to ongoing (Euro)WordNet initiatives.

We are currently investigating the use of the lexicon for word-sense disambiguation in machine-translation

[5]Ultimately, the misclassed verbs should be disassociated from the HowNet concept, but there is currently no way to tease apart such cases from the Optilex gaps. Thus, the two are treated identically.

and cross-language information retrieval. As we saw above the Chinese verb 拉 (la) has several possible translations, but not all of these will be appropriate in every context. If we can determine which HowNet concept corresponds to 拉 (la), then we will translate it appropriately. For example, if the HowNet concept is |Transport|, the translation would be *ship* or *transport*, but not *slash*, *chat*, *implicate*, etc. We can detect which HowNet class is appropriate by examining the other words in the sentence. If those words co-occur with *other* Chinese verbs associated with a particular HowNet concept (as determined through a corpus analysis), then it is likely that that HowNet concept is the appropriate one for the Chinese verb. That is, if we find other verbs from a given HowNet concept occurring in the same context, then we can hypothesize that this particular verb has the meaning of this HowNet concept.

The algorithm for mapping between HowNet concepts and EVCA classes requires a "training" step—i.e., the seed mappings given earlier. However, it is possible to produce a ranked mapping between semantic-role specifications by counting correspondences between EVCA-based roles and the HowNet-based roles across the entire concept space. This approach is also currently under investigation.

Another area of investigation is the use of a WordNet-based distance metric (e.g., the information-content approach of (Resnik, 1995)) for additional pruning power in the HowNet-to-EVCA alignment. Because each of the entries in the EVCA classification is associated with a WordNet sense, it is possible to rule out certain class assignments for a given HowNet concept by examining semantic distance between the Optilex glosses for a particular Chinese word and the glosses for other words associated with that concept.

References

Hoa Trang Dang, Karin Kipper, Martha Palmer, and Joseph Rosenzweig. 1998. Investigating Regular Sense Extensions Based on Intersective Levin. In *ACL/COLING 98, Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics* (joint with the 17th International Conference on Computational Linguistics), pages 293–306, Montreal, Canada, August 10-14.

Bonnie J. Dorr and Douglas Jones. 1996. Acquisition of Semantic Lexicons: Using Word Sense Disambiguation to Improve Precision. In *Proceedings of the Workshop on Breadth and Depth of Semantic Lexicons, 34th Annual Conference of the Association for Computational Linguistics*, pages 42–50, Santa Cruz, CA.

Bonnie J. Dorr and Douglas Jones. 1999. Acquisition of semantic lexicons: Using word sense disambiguation to improve precision. In Evelyne Viegas, editor, *Breadth and Depth of Semantic Lexicons*. Kluwer Academic Publishers, Norwell, MA.

Bonnie J. Dorr, Nizar Habash, and David Traum. 1998. A Thematic Hierarchy for Efficient Generation from Lexical-Conceptal Structure. In *Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA-98, in Lecture Notes in Artificial Intelligence, 1529*, pages 333–343, Langhorne, PA, October 28-31.

Bonnie J. Dorr. 1997. Large-Scale Acquisition of LCS-Based Lexicons for Foreign Language Tutoring. In *Proceedings of the ACL Fifth Conference on Applied Natural Language Processing (ANLP)*, pages 139–146, Washington, DC.

Douglas Jones, Robert Berwick, Franklin Cho, Zeeshan Khan, Karen Kohl, Naoyuki Nomura, Anand Radhakrishnan, Ulrich Sauerland, and Brian Ulicny. 1994. Verb Classes and Alternations in Bangla, German, English, and Korean. Technical report, Massachusetts Institute of Technology.

Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation.* University of Chicago Press, Chicago, IL.

George A. Miller and Christiane Fellbaum. 1991. Semantic Networks of English. In Beth Levin and Steven Pinker, editors, *Lexical and Conceptual Semantics, Cognition Special Issue*, pages 197–229. Elsevier Science Publishers, B.V., Amsterdam, The Netherlands.

Naoyuki Nomura, Douglas A. Jones, and Robert C. Berwick. 1994. An architecture for a universal lexicon: A case study on shared syntactic information in Japanese, Hindi, Ben Gali, Greek, and English. In *Proceedings of COLING-94*, pages 243–249.

Mari Broman Olsen, Bonnie J. Dorr, and Scott C. Thomas. 1998. Enhancing Automatic Acquisition of Thematic Structure in a Large-Scale Lexicon for Mandarin Chinese. In *Proceedings of the Third Conference of the Association for Machine Translation in the Americas, AMTA-98, in Lecture Notes in Artificial Intelligence, 1529*, pages 41–50, Langhorne, PA, October 28-31.

Martha Palmer and Joseph Rosenzweig. 1996. Capturing motion verb generalizations with synchronous tags. In *Proceedings of the Second Conference of the Association for Machine Translation in the Americas*, Montreal, Quebec, Canada.

Martha Palmer and Zhibao Wu. 1995. Verb Semantics for English-Chinese Translation. *Machine Translation*, 10(1–2):59–92.

P. Procter. 1978. *Longman Dictionary of Contemporary English.* Longman, London.

Philip Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of IJCAI-95*, pages 448–453, Montreal, Canada.

Patrick Saint-Dizier. 1996. Semantic Verb Classes Based on 'Alternations' and on WordNet-like Semantic Criteria: A Powerful Convergence. In *Proceedings of the Workshop on Predicative Forms in Natural Language and Lexical Knowledge Bases*, pages 62–70, Toulouse, France.

Dong Zhendong. 1988a. Enlightment and Challenge of Machine Translation. *Shanghai Journal of Translators for Science and Technology*, 1:9–15.

Dong Zhendong. 1988b. Knowledge Description: What, How and Who? In *Proceedings of International Symposium on Electronic Dictionary*, page 18, Tokyo, Japan.

Dong Zhendong. 1988c. MT Research in China. In *Proceedings of International Conference on New Directions in Machine Translation*, pages 85–91, Budapest. Also in New Directions in Machine Translation, 4 Distributed Language Translation edited by Dan Maxwell, Klaus Schubert and Toon Witkam, Foris Publications, Dordrecht.